

Availability of Federal Data: Policy Considerations for Disclosure, Preservation, and Governance

March 31, 2026

Congressional Research Service

<https://crsreports.congress.gov>

R48889



R48889

March 31, 2026

Meghan M. Stuessy
Analyst in Government
Organization and
Management

Taylor R. Knoedl
Analyst in American
National Government

Availability of Federal Data: Policy Considerations for Disclosure, Preservation, and Governance

Federal data can provide valuable information for various audiences—from farmers seeking to protect bats that eat crop-harming insects to local efforts determining where to rebuild to avoid coastal flooding. In 2013, the Office of Management and Budget (OMB) described openly available federal data and statistical information as “a valuable national resource and strategic asset” that, when made accessible, discoverable, and usable by the public, “can help fuel entrepreneurship, innovation, and scientific discovery.”

Efforts to make federal data more readily available have evolved over time. Such data may have been stored and filed in hard and paper copies and later in software and electronic formats. Today, certain data may be retrieved through agency websites or on Data.gov. Data.gov itself is a case study for open data, intended to demonstrate that making federal data available can help agencies avoid duplicative internal research, enable the discovery of complementary datasets held by other agencies, and empower employees to make better-informed, data-driven decisions, among other benefits.

Throughout 2025, media reports have suggested that the availability of federal data has been reduced. Some observers are also tracking the removal of specific datasets, variables, and tools. In parallel, changing public perspectives on data availability may demand new levels of data access, such as making data available for predictable periods of time, in a variety of software-compatible formats, and with appropriate descriptive metadata for easing findability and usability of the information. While statute discusses when and how information is to be added to Data.gov, it does not explain whether and how information may be removed. Although researchers and the public may derive value from being able to trace data over time to determine changes in trends or collection methods, the statute does not explicitly consider versioning requirements for agency data. However, requiring these attributes for Data.gov may help address or clarify difficulties in measuring data availability. Congress may be interested in determining whether there are trends to certain data becoming available or to when data is altered and removed. Such trends may provide insight and direction for Congress to further examine agency activities or make decisions to support new data use cases.

Information availability, of which data availability is a type, can be considered the intersection of when and how information is released. Section 3542 of Title 44 of the *U.S. Code* defines *information availability* as “ensuring timely and reliable access to and use of information.” Generally, statute and associated OMB guidance contemplates two types of information availability in terms of timing: (1) proactive disclosure and information dissemination and (2) request-based disclosure. Certain types of data have specific requirements in terms of formatting and structure to ensure that the information can be made available and potentially archived.

This report examines the variables of federal data availability and its policy underpinnings. The report discusses the state and concept of federal data availability and explains the information life cycle framework. It explains how information may be made available proactively or upon request through existing mechanisms and also explains statutory requirements for information dissemination, preservation, and whether and when information can be removed. The report concludes with policy options for Congress, including a review of efforts to preserve federal data through web captures; examining controls to assess data versioning, sourcing, and modifications; and, finally, considerations for implementing data governance and transparency mechanisms throughout agency structures.

Contents

What Is Data Availability?.....	2
Public Information: Agency Proactive Disclosure Requirements.....	5
The Freedom of Information Act and Proactive Disclosure	5
The Paperwork Reduction Act	6
The OPEN Government Data Act.....	8
Requesting Information: Agency Disclosure Procedures.....	10
Requesting Information via the Freedom of Information Act.....	10
Requesting Statistical Data via the Standard Application Process.....	11
Preserving and Disposing of Data	13
The Federal Records Act and Records Preservation	13
Maintaining and Disposing of Federal Records	14
Technical Details Required to Archive Data	15
Removing Data.....	16
Corrections Under the Information Quality Act	17
Removal of Convenience Copies and Web Records.....	18
Measuring Data Availability	19
Independent Efforts to Archive Data	22
Issues for Congress.....	23
Web Capture.....	24
Data Fixity and Provenance	24
Data Governance and Transparency.....	25

Figures

Figure 1. The Information Life Cycle	4
--	---

Appendixes

Appendix. Glossary	27
--------------------------	----

Contacts

Author Information.....	28
-------------------------	----

Policymakers, members of the public, and businesses regularly use federal data and statistical information to make policy choices, programmatic investments, and funding decisions. Although nonfederal entities may produce data, users often prefer federal data for its credibility, accuracy, and objectivity.¹ Federal data can provide valuable information for various audiences—from farmers seeking to protect bats that eat crop-harming insects to local efforts determining where to rebuild to avoid coastal flooding.² In 2013, the Office of Management and Budget (OMB) described openly available federal data and statistical information as “a valuable national resource and strategic asset” that, when made accessible, discoverable, and usable by the public, “can help fuel entrepreneurship, innovation, and scientific discovery.”³ However, making data available also requires the investment of time, resources, and personnel and may compete with other agency priorities.

Efforts to make federal data more readily available have evolved over time: Originally, such data may have been stored and provided in paper and other hard copy formats (e.g., photo copies, film, and microfiche) and later stored in hard copy and locally maintained digital formats but provided only in hard copy formats (e.g., floppy discs, CDs). Today, data is increasingly stored digitally and provided electronically—with certain digital data available on-demand from anywhere with an internet connection through agency websites. Nonetheless, such efforts to make federal data more available have been implemented in piecemeal fashion, and certain information now remains in both hard copy and electronic formats and may be stored locally or remotely in the cloud.⁴

Meanwhile, public expectations are likely to continue shifting toward assuming complete, customizable, and on-demand digital data availability. Agency responses will likely shift to meet those expectations and need to adapt to the requirements of digital data storage and delivery. For example, the public may have previously expected a published print volume once a year and now they may expect more frequent and current digital data to be available online. For agencies, serving the public may mean establishing new expectations through communicating new schedules for digital publishing, providing for data downloads compatible with a variety of software applications, and including appropriate descriptive metadata alongside the digital data for easing findability and usability of the data.

In 2018, Congress enacted the multi-titled Foundations for Evidence-Based Policymaking Act of 2018 (FEBPA, P.L. 115-435), which established certain agency roles, practices, and requirements related to agency use of information in learning agendas, the cataloging and dissemination of agency data, and statistical and confidentiality protocols, among other things. As part of the implementation of FEBPA, on January 15, 2025, OMB issued guidance to agencies on how to comply with the law’s data availability, data management, and dissemination requirements.⁵ At

¹ American Statistical Association, “The Nation’s Data at a Crossroads: Year Two Status Report,” September 3, 2025, pp. 5-8, <https://www.amstat.org/docs/default-source/amstat-documents/nations-data-at-crossroads.pdf>. See also American Statistical Association, “The Nation’s Data at Risk: Meeting America’s Information Needs for the 21st Century,” July 9, 2024, p. 2, <https://www.amstat.org/docs/default-source/amstat-documents/the-nation's-data-at-risk-supporting-materials/valueoffederalstatistics.pdf>.

² For a range of federal data use cases, see America’s Essential Data, “Use Case Repository,” <https://essentialdata.us/use-cases.html>. See also Nicholas Eberstadt et al., “‘In Order That They Might Rest Their Arguments on Facts’: The Vital Role of Government-Collected Data,” American Enterprise Institute, March 2017.

³ OMB, “Open Data Policy—Managing Information as an Asset,” M-13-13, May 9, 2013, p. 1, <https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>.

⁴ See also CRS In Focus IF12899, *Data Centers and Cloud Computing: Information Technology Infrastructure for Artificial Intelligence*, by Ling Zhu.

⁵ OMB, “Phase 2 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Open (continued...) ”

the same time, some stakeholders have raised concerns about the amount of federal data that remains available. For example, in late January 2025, several news outlets reported the apparent removal of datasets from Data.gov.⁶ These developments may raise questions for Congress in terms of (1) whether and how current law sufficiently captures policymakers' intent for making data available via existing information management and data governance structures and (2) what challenges agencies and the public face in locating and assessing existing data resources for their own use cases.

This report examines the different dimensions of federal data availability: When is data available and to which audiences? Is the data genuine and unaltered? Has the data been corrected, and when should it be removed from active use? The report discusses the state and concept of federal data availability using the information life cycle framework. The report explains how information, including data, may be made available proactively or upon request. It also explains statutory requirements for information dissemination, preservation, and whether and when information and data can be removed. The report concludes with policy options for Congress.

What Is Data Availability?

Scholars have long debated the contours of *data* and *information* and whether they are discrete or overlapping concepts, leading to variations in how they are defined.⁷ Although the terms *data* and *information* can commonly be used interchangeably, in the federal government context, *data* may be considered to be “a value or set of values representing a specific concept or concepts.”⁸ In contrast, *information* may be defined as “any communication or representation of knowledge such as facts, data, or opinions in any medium or form, including textual, numerical, graphic, cartographic, narrative, electronic, or audiovisual forms.”⁹ Viewing the two definitions in tandem, therefore, *data* is a specific type of *information*.

In an effort to further organize the concept of *data*, both the National Institute of Standards and Technology (NIST) and OMB define *data* by its form of structure.¹⁰ *Structured data*, for example, may be found in a database that clearly indicates what type of information each field contains, or

Government Data Access and Management Guidance,” M-25-05, January 15, 2025, p. 1, <https://bidenwhitehouse.archives.gov/wp-content/uploads/2025/01/M-25-05-Phase-2-Implementation-of-the-Foundations-for-Evidence-Based-Policymaking-Act-of-2018-Open-Government-Data-Access-and-Management-Guidance.pdf>.

⁶ Mike Stobbe and Mike Schneider, “Trump Administration’s Data Deletions Set Off ‘a Mad Scramble,’ Researcher Says,” Associated Press, February 3, 2025, <https://apnews.com/article/cdc-census-federal-data-trump-6a9ba7c01a42b72e2c0a119325ba3753>; and Cynthia Cox et al., “A Look at Federal Health Data Taken Offline,” KFF, February 2, 2025, <https://www.kff.org/policy-watch/a-look-at-federal-health-data-taken-offline/>.

⁷ Michael K. Buckland, “Information as Thing,” *Journal of the American Society for Information Science*, vol. 42, no. 5 (June 1991), pp. 311-388.

⁸ General Services Administration (GSA), “Glossary: Data vs. Information,” <https://resources.data.gov/glossary/data-vs.-information/>.

⁹ OMB, *Managing Information as a Strategic Resource*, Circular A-130, July 2016, p. 29, https://www.whitehouse.gov/wp-content/uploads/legacy_drupal_files/omb/circulars/A130/a130revised.pdf.

¹⁰ William Newhouse et al., *Data Classification Concepts and Considerations for Improving Data Protection*, National Institute of Standards and Technology, November 2023, p. 3, <https://nvlpubs.nist.gov/nistpubs/ir/2023/NIST.IR.8496.ipd.pdf>; and OMB, “Phase 2 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Open Government Data Access and Management Guidance,” M-25-05, January 15, 2025, p. 6, <https://bidenwhitehouse.archives.gov/wp-content/uploads/2025/01/M-25-05-Phase-2-Implementation-of-the-Foundations-for-Evidence-Based-Policymaking-Act-of-2018-Open-Government-Data-Access-and-Management-Guidance.pdf>.

rectangular or tabular data organized into rows and columns.¹¹ On the other hand, information that is considered to be *unstructured*—such as documents, pictures, audio, or video—do not follow a specific format and could contain nearly any type of information.

Certain statutory definitions and institutional preferences discussed later in this report may also appear to weigh in on this terminology issue, with some laws and policies discussing information management, which subsumes the discussion of data within those policies, and others, as is the case in the OPEN Government Data Act, using the term *data* to discuss a particular type of information. Similarly, for purposes of this report, the term *data* refers to the type of structured information organized in a standardized format, and *data* will refer to a specific subset of *information*. *Information*, in this report, comprises the larger landscape of textual, graphical, opinion, data, and narrative communications. The regulations and practices discussed in this report that govern *information* also apply to *data*.

The idea of data availability is comprised of various dimensions, each with a slightly different implication for how information or data is published, shared, or stored. For example, *findability*—the concept that the information can be located and discovered—may be one dimension. Another is *reusability*—the concept of information being archived and well-described that enables product replication or use in different settings.¹² Although these are just two perspectives, decisions on what dimension of availability to prioritize has impacts on the statutory and policy landscape supporting and directing federal information availability.

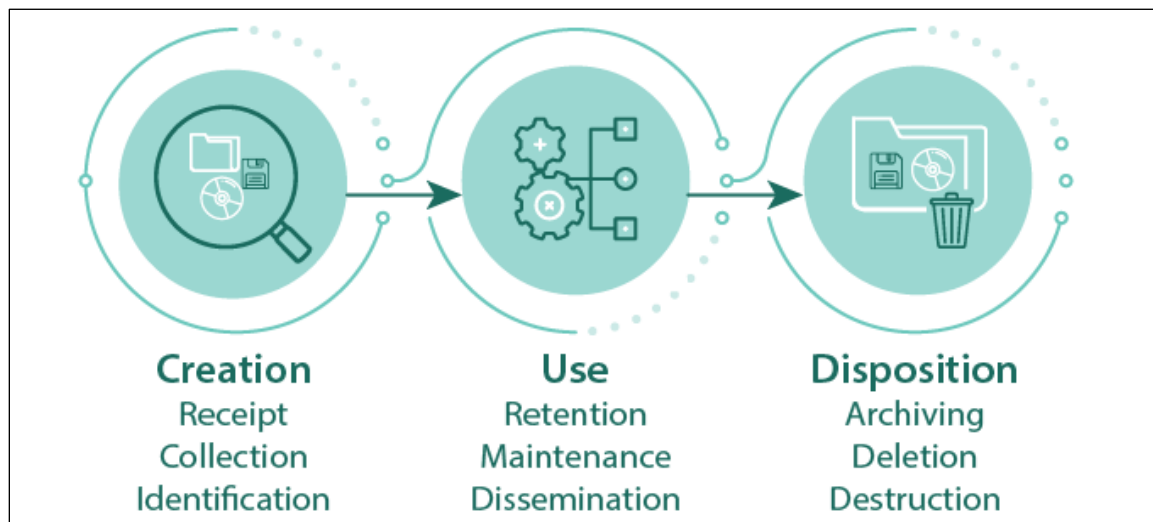
Decisions about when and how to make data available vary by agency and over what can be thought of as the life of the data. For example, findability may be more of an issue to users once the information has moved from regular use to storage. Reusability may be a user priority in the event new information cannot be obtained due to limited resources. The dimensions of availability may vary in importance as data flows among users, purposes, and systems.

OMB and the National Archives and Records Administration (NARA) use the concept of the information life cycle to better understand, direct, and manage the flow of data. The information life cycle consists of three stages where the information is (1) collected or created, (2) used or shared, and (3) stored or disposed of. Data storage requirements have direct impacts on how information is created and used within agencies.¹³

¹¹ OMB, “Phase 2 Implementation,” p. 6.

¹² Mark D. Wilkinson et al., “The FAIR Guiding Principles for Scientific Data Management and Stewardship,” *Scientific Data*, vol. 3 (March 15, 2016), p. 4, <https://www.nature.com/articles/sdata201618>.

¹³ OMB, *Managing Information*, Circular A-130, p. 29. See also NARA, “What’s a Record?,” August 15, 2016, <https://www.archives.gov/about/info/whats-a-record.html>.

Figure 1. The Information Life Cycle

Source: CRS depiction of Office of Management and Budget and National Archives and Records Administration definition of information life cycle.

At all three stages of the information life cycle, one may consider different responses to when and how the information should be made available. For example, information may be available within the creating agency while the agency is implementing a particular program, but it could be more broadly available to the public at later stages in the information life cycle. *Information availability* may also refer to the format of the information and whether the format is consistently able to be used. For example, information available in hard copy and formatted for a particular software (such as a Blu-ray disc recording being unviewable on a standard DVD player) might be useful only to recipients of the media with the corresponding software, while information on a website might be more broadly accessible. Each of these choices presents particular benefits and risks, where certain decisions may be better suited to specific use cases.

Information availability, then, can be considered the intersection of when and how information is released.¹⁴ Statute defines *information availability* as “ensuring timely and reliable access to and use of information,” blending together when and how information is released into one concept.¹⁵ Generally, statute and associated OMB guidance contemplates two types of information availability in terms of user access: (1) proactive disclosure and information dissemination, where information of general interest is publicly available, and (2) request-based disclosure, where users can request information of particular interest, sometimes in specific formats. Certain types of data have administrative requirements in terms of formatting and structure to ensure how the information can be made available and potentially archived.¹⁶

¹⁴ Kaye Timonera, “What Is Data Availability? Best Practices and Challenges,” Datamation, March 1, 2024, <https://www.datamation.com/big-data/data-availability/>.

¹⁵ 44 U.S.C. §3542.

¹⁶ 36 C.F.R. §1235 Subpart C.

Public Information: Agency Proactive Disclosure Requirements

Proactive disclosure refers to a type of information availability where agencies make records and data publicly available without waiting for specific requests for the information.¹⁷ Proactive disclosure of certain general information by agencies is required by the Freedom of Information Act, the Paperwork Reduction Act, and the OPEN Government Data Act. This section focuses on government-wide statutes requiring proactive disclosure of information, although other types of information may be required to be made proactively available by individual or agency-specific statutes.

The Freedom of Information Act and Proactive Disclosure

The Freedom of Information Act (FOIA, 5 U.S.C. §552) is often referred to as the embodiment of “the people’s right to know” about the activities and operations of government, and it was enacted in response to Congress’s perception of improper secrecy in the executive branch.¹⁸ It established a statutory presumption of public access to information held by executive branch agencies. FOIA generally allows any person—individual or corporate, U.S. citizen or not—to request and obtain, without explanation or justification, a large variety of records and information held by the federal government.

FOIA establishes a presumption that the public should have access to information in the possession of executive branch agencies and departments of the federal government. Prior to FOIA, a requester had the burden of proof to show a “need to know” to gain access to government information. FOIA assumes a “right to know” and shifts the burden to the federal government to establish a need to keep the information secret.¹⁹ FOIA establishes a three-part system that requires federal agencies to disclose government information to the public.

First, FOIA directs agencies to publish substantive and procedural rules, along with certain other important government materials, in the *Federal Register*.²⁰ Second, agencies must electronically disclose a separate set of information that consists of, among other things, final adjudicative opinions and certain “frequently requested” records.²¹ With respect to information availability, the Department of Justice (DOJ) explains what it means to electronically publish information and how agencies should identify frequently requested materials. The third, discussed later in this report, allows the public to request certain information from agencies.

Since FOIA’s enactment in 1966, agencies made records “available for public inspection and copying” by placing hard copies of the records in government “Reading Rooms.”²² In 1996, Congress amended FOIA to clarify that agencies shall make subject materials “available for

¹⁷ Department of Justice (DOJ), “Department of Justice Guide to the Freedom of Information Act: Proactive Disclosures,” March 14, 2025, p. 1, https://www.justice.gov/oip/foia-guide/proactive_disclosures/dl?inline.

¹⁸ Congress sought to eliminate the burden of proof that had existed under the public information section of the Administrative Procedure Act, which required requesters to establish a justification or a need for information being sought (P.L. 79-404, §3 [1946]). Under FOIA, in contrast, public access is presumed, and federal agencies must justify denying access to requested information.

¹⁹ U.S. Congress, House Committee on Oversight and Government Reform, *A Citizen’s Guide to Using the Freedom of Information Act and the Privacy Act to Request Government Records*, committee print, 112th Cong., 2nd sess., September 21, 2012, H.Rept. 112-689, p. 2, <https://www.congress.gov/112/crpt/hrpt689/CRPT-112hrpt689.pdf>.

²⁰ 5 U.S.C. §552(a)(1).

²¹ 5 U.S.C. §552(a)(2).

²² DOJ, “Proactive Disclosure of Non-Exempt Agency Information: Making Information Available Without the Need to File a FOIA Request,” October 26, 2022, https://www.justice.gov/oip/oip-guidance/proactive_disclosure_of_non-exempt_information.

public inspection in an electronic format.”²³ At the time, Members suggested that the bill “would encourage agencies to offer online access to Government information, effectively transforming an individual’s home computer into a Government agency’s public reading room.”²⁴ The accompanying House Committee on Government Reform and Oversight report similarly noted that, like information published in the *Federal Register*, agency information and frequently requested records should be made available online.²⁵

FOIA requires agencies to make available in an electronic format records that “have become or are likely to become the subject of subsequent requests for substantially the same records” or those that have been requested three or more times.²⁶ The House report notes that this requirement “would help to reduce the number of multiple FOIA requests for the same records requiring separate agency responses” and enable agencies to use their resources to respond to new and unique requests.²⁷ Echoing this sentiment, DOJ explains that proactive disclosures “efficiently satisfy the demand for records of interest to multiple people” and provides agencies with a checklist on how to implement FOIA’s proactive disclosure requirements.²⁸

The Paperwork Reduction Act

The Paperwork Reduction Act of 1980 required agencies to conduct certain information resources management activities and further required each agency head to designate a “senior official” to carry out responsibilities related to the coordination of federal information policy.²⁹ Building upon these requirements, the Paperwork Reduction Act of 1995 stipulated categories of information resources management responsibilities for agencies, including the categories of (1) information dissemination and (2) statistical policy and coordination.³⁰ These categories relate to existing proactive disclosure requirements by further requiring agencies to distribute certain agency information and describe agency procedures and findings to the public.

²³ 5 U.S.C. §552(2).

²⁴ “Electronic Freedom of Information Act Amendments of 1996,” House debate, *Congressional Record*, vol. 142, part 128 (September 17, 1996), p. H10451.

²⁵ U.S. Congress, House Oversight and Government Reform Committee, *Electronic Freedom of Information Amendments of 1996*, To accompany H.R. 3802, 104th Cong., 2nd sess., September 17, 1996, H.Rept. 104-795, pp. 20-21. See also 44 U.S.C. §4101.

²⁶ 5 U.S.C. §552(a)(2)(D).

²⁷ H.Rept. 104-795, p. 21.

²⁸ DOJ, “Implementation Checklist for OIP Guidance on Proactive Disclosures of Non-Exempt Agency Information,” December 5, 2022, <https://www.justice.gov/oip/implementation-checklist-oip-guidance-proactive-disclosures-non-exempt-agency-information>.

²⁹ P.L. 96-511, 94 Stat. 2819. Under the Clinger-Cohen Act of 1996, the terminology of “senior official” was redesignated as the “Chief Information Officer.” P.L. 104-106, §5125, 110 Stat. 684. This provision was originally enacted as part of the Information Technology Management Reform Act of 1996, in Division E of P.L. 104-106 (110 Stat. 679), the National Defense Authorization Act for Fiscal Year 1996 (110 Stat. 186). Subsequently, Section 808 of P.L. 104-208 (110 Stat. 3009-393) retitled Divisions D (the Federal Acquisition Reform Act of 1996, 110 Stat. 642) and E as the Clinger-Cohen Act of 1996. For more information about chief information officers, see CRS Report R48147, *Chief Information Officers (CIOs): Agency Roles and Responsibilities*, by Meghan M. Stuessy and Dominick A. Fiorentino.

³⁰ See also “Information Resources Management (IRM)” in CRS Report R48147, *Chief Information Officers (CIOs): Agency Roles and Responsibilities*, by Meghan M. Stuessy and Dominick A. Fiorentino.

Agency Requirements to Disseminate Information

Agencies are required to conduct certain information dissemination activities by statute in accordance with guidance from the OMB director. Among other requirements, under Title 44, Section 3506(d), of the *U.S. Code*, an agency shall:

- regularly solicit and consider public input on the agency’s information dissemination activities;
- ensure that the public has timely and equitable access to the agency’s public information; and
- provide adequate notice when initiating, substantially modifying, or terminating significant information dissemination products.

The accompanying Senate report explains, “These provisions require OMB to develop government-wide policies and guidelines for information dissemination and to promote public access to information maintained by Federal agencies.”³¹ However, specific guidance related to information dissemination appears to be limited to OMB Memorandum M-06-02, published in December 2005.³² The Senate report further describes that “OMB has an obligation to promote public access to government information through the development and oversight of government-wide information dissemination policies. Likewise, agencies have an obligation to conduct their dissemination activities to ensure that the public has timely and equitable access to public information.”³³ Congress may wish to consider whether OMB guidance adequately promotes and reinforces agency requirements with respect to information dissemination.

Agency Requirements to Coordinate and Provide Statistical Information³⁴

Federal agencies are required in statute to provide quality and transparency standards on information collected for statistical purposes. Such information is generally made available to the public after disclosure avoidance protections are applied.³⁵ Section 3506(e) of Title 44 of the *U.S. Code* states that each agency as it pertains to statistical policy and coordination shall:

- ensure the relevance, accuracy, timeliness, integrity, and objectivity of information collected or created for statistical purposes;
- inform respondents fully and accurately about the sponsors, purposes, and uses of statistical surveys and studies;

³¹ U.S. Congress, Senate Governmental Affairs Committee, To further the goals of the Paperwork Reduction Act to have federal agencies become more responsible and publicly accountable for reducing the burden of federal paperwork on the public and for other purposes, Report to accompany S. 244, 104th Cong., 1st sess., February 14, 1995, S.Rept. 104-8, p. 3.

³² OMB Memorandum M-06-02’s full subject is “Improving Public Access to and Dissemination of Government Information and Using the Federal Enterprise Architecture Data Reference Model.” See also OMB, *Managing Information as a Strategic Resource*, Circular A-130, p. 3. However, OMB guidance on both information dissemination and statistical policy and coordination is not in a single location. Last revised in 2016, Circular A-130 states, “Although this Circular touches on many specific information resources management issues such as privacy, confidentiality, information quality, dissemination, and statistical policy, those topics are covered more fully in other Office of Management and Budget (OMB) policies, which are available on the OMB website.”

³³ S.Rept. 104-8, p. 25.

³⁴ This section was authored by Taylor R. Knoedl, Analyst in American National Government.

³⁵ For more information, see Census Bureau, “Statistical Safeguards,” January 22, 2026, https://www.census.gov/about/policies/privacy/statistical_safeguards.html.

- protect respondents' privacy and ensure that disclosure policies fully honor pledges of confidentiality;
- observe federal standards and practices for data collection, analysis, documentation, sharing, and dissemination of information;
- ensure the timely publication of the results of statistical surveys and studies, including information about the quality and limitations of the surveys and studies; and
- make data available to statistical agencies and readily accessible to the public.

Title 44 also provides the basis for several governance elements of the federal statistical system, including the establishment of the chief statistician role housed within OMB's Office of Information and Regulatory Affairs and the Interagency Council on Statistical Policy (ICSP), which is chaired by the chief statistician.³⁶

The chief statistician has the broad responsibility of providing coordination, guidance, and oversight of the federal statistical agencies and their activities. Coordination of the federal statistical system includes the development of Statistical Policy Directives (SPDs) which are generally authorized under Section 3504(e) of Title 44 of the *U.S. Code*. These directives are issued as needed to ensure the quality and coordination of federal statistical activities.³⁷

Access and dissemination of federal statistical products is required through SPD No. 4: Release and Dissemination of Statistical Products Produced by Federal Statistical Agencies. Under SPD No. 4, "statistical agencies must ensure that all users have equitable and timely access to data that are disseminated to the public." Furthermore, SPD No. 4 allows federal statistical agencies to issue their statistical products in printed and/or electronic formats. They are required to publish statistical products on their websites.³⁸

The OPEN Government Data Act

The Open, Public, Electronic, and Necessary Government Data Act—also called the OPEN Government Data Act (Title II of FEBPA, P.L. 115-435)—seeks to change how government information is formatted, catalogued, and presented for public access and use. The law expands on the requirement in FOIA for agencies to make electronic copies of previously released records more broadly available for public inspection. OMB issued related guidance for the act on January 15, 2025, in the form of OMB Memorandum M-25-05.³⁹ In this way, the OPEN Government Data Act and its implementation serve to create another form of proactive disclosure.

³⁶ 44 U.S.C. §3504(e)(8). For more information about the federal statistical system, see CRS Report R48161, *The Federal Statistical System: An Overview*, by Taylor R. Knoedl.

³⁷ Bureau of Labor Statistics, "Statistical Policy Directives of the Office of Management and Budget," <https://www.bls.gov/bls/statistical-policy-directives.htm>.

³⁸ OMB, "Statistical Policy Directive No. 4: Release and Dissemination of Statistical Products Produced by Federal Statistical Agencies," 73 *Federal Register* 12622, March 7, 2008, <https://www.govinfo.gov/content/pkg/FR-2008-03-07/pdf/E8-4570.pdf>.

³⁹ OMB, "Phase 2 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Open Government Data Access and Management Guidance," M-25-05, January 15, 2025, <https://bidenwhitehouse.archives.gov/wp-content/uploads/2025/01/M-25-05-Phase-2-Implementation-of-the-Foundations-for-Evidence-Based-Policymaking-Act-of-2018-Open-Government-Data-Access-and-Management-Guidance.pdf>. In September 2021, OMB staff told the Government Accountability Office (GAO) that the delay in finalizing the guidance was "due to delays resulting from the Coronavirus Disease 2019 (COVID-19) pandemic and the transition to a new presidential administration." For more information, see GAO, *Open Data: Additional Action* (continued...)

The law generally requires that agency data assets be inventoried, formatted, and presented for the public's access and use through a federal data catalogue.⁴⁰ While it is not explicitly stated in statute, it appears that role of Data.gov is to be the current tool fulfilling this catalogue requirement. OMB acknowledges that the federal data catalogue is currently accessible through Data.gov "or any successor website."⁴¹ Statute defines a *data asset* as "a collection of data elements or data sets that may be grouped together," and M-25-05 further describes data assets as "composed of structured or semi-structured data," such as tabular data organized into rows and columns or a database of digital images, and "logically grouped," where the data is grouped together based on similar characteristics, a shared function or purpose, or some other logical method.⁴² Further, M-25-05 clarifies that the definition is intentionally broad: "a database procured through a contract may be a data asset subject to the requirements of this guidance, even if the contents of the database are owned by a private party."⁴³

Consistent with statute, M-25-05 requires each agency to develop and maintain a comprehensive data inventory (CDI) "that accounts for all data assets created by, collected by, under the control or direction of, or maintained by the agency."⁴⁴ Although a CDI is to assess and categorize all of an agency's data assets, both law and guidance stipulate that a data asset's itemization does not necessarily mean that the data asset can be made publicly available. OMB explains that the federal data catalogue, located at Data.gov, "will be a central source for the public and other agencies to discover agency data assets that have not been and may never be disseminated." The federal data catalogue may also serve as a resource for other agencies seeking to leverage or reuse existing resources.

While agency CDIs may promote the disclosure of certain data assets by making the public aware of them, agencies are permitted to redact portions of their CDIs consistent with FOIA's exceptions.⁴⁵ The CDIs are required to provide specific descriptive metadata elements for each data asset, including the name and description of the data asset; the names and definitions of the included variables; the date it was added to the CDI; when it was posted and/or updated; and security and privacy categorizations describing the availability, use, and levels of access to the data asset.⁴⁶

M-25-05 requires agency CDIs to supply certain metadata about the data assets, such as a description, title, and names and definitions of data asset variables, as well as additional metadata that conforms with the DCAT-US 3.0 standard.⁴⁷ This requirement follows from recommendations from the Federal Chief Data Officers Council and the Federal Committee on Statistical Methodology to make federal data more discoverable and usable.⁴⁸ The recommendation seeks to make U.S. data cataloguing standards current with international

Required for Full Public Access, GAO-22-104574, December 2021, p. 12, <https://www.gao.gov/assets/gao-22-104574.pdf>.

⁴⁰ 44 U.S.C. §3511.

⁴¹ OMB, "Phase 2 Implementation," p. 17.

⁴² OMB, "Phase 2 Implementation," pp. 6-7.

⁴³ OMB, "Phase 2 Implementation," p. 7.

⁴⁴ 44 U.S.C. §3511(a)(1).

⁴⁵ OMB, "Phase 2 Implementation," p. 10; OMB, "Open Data Policy—Managing Information as an Asset," p. 1.

⁴⁶ 44 U.S.C. §3511(a)(2)(A).

⁴⁷ OMB, "Phase 2 Implementation," p. 8. See also Thomas Dabolt et al., "DCAT-US—Version 3: Data Catalog Application Profile for the United States of America," May 25, 2025, <https://doi-do.github.io/dcat-us/>.

⁴⁸ Madison Alder, "White House Nearing Finish Line on Federal Data Guidance," *Fedscoop*, October 2024, <https://fedscoop.com/white-house-federal-data-guidance-nearing-finish/>.

standards and enable data providers to “accurately and efficiently document federal data for public and internal re-use.”⁴⁹ Further, the federal data catalogue “does not host or maintain any data assets directly; rather it provides a centralized point of entry to discover government data assets.”⁵⁰ Because data assets remain hosted by the agency, agency records management programs and policies govern the underlying availability of the federal data.

Requesting Information: Agency Disclosure Procedures

Individuals may request access to information or request that certain types of information be disclosed. Under FOIA, any member of the public may request existing government information from covered federal agencies, subject to nine statutory disclosure exemptions. For statistical data, researchers are able to request information for specified projects through a standard application process (SAP). In each, however, the request for information does not necessarily guarantee complete access and does not require agencies to comply with providing the information at specific levels of granularity (e.g., data cuts by levels of government or by other individual variables) or in particular formats.

Requesting Information via the Freedom of Information Act

Following FOIA’s two tenets concerning proactive disclosure of agency information, the third part is FOIA’s request-based system of disclosure. Any member of the public may request existing government information from covered federal agencies. A request must reasonably describe the records sought and be made in accordance with an agency’s published rules.⁵¹ However, FOIA’s presumptive right of access is limited when the requested information falls within the scope of nine statutory exemptions.⁵² Under these exemptions, agencies may withhold information that, for example, relates to national defense or foreign policy, trade secrets, or certain law enforcement records.⁵³

FOIA also permits the public to request agency records “in any format, including an electronic format,” emphasizing that the released information should also be usable.⁵⁴ As part of the 1996 amendments to FOIA, the law requires that an agency “shall provide the record in any form or format requested by the person if the record is readily reproducible by the agency in that form or format. Each agency shall make reasonable efforts to maintain its records in forms or formats that are reproducible for purposes of this section.”⁵⁵ DOJ echoes this, explaining, “Beyond the legal requirements imposed by the FOIA, agencies should, as a matter of discretion, be routinely posting material that is of interest to the public, taking advantage of technology and new tools to make that posted data usable and easily accessible.”⁵⁶

⁴⁹ Thomas Dabolt and Michael Ratcliff, “CDOC Completion of the Building Trust and FAIRness into the Process for Finding and Using Government Data Project (FAIRness Project),” Federal Chief Data Officers Council, September 27, 2024, <https://www.cdo.gov/fairness-project/>.

⁵⁰ OMB, “Phase 2 Implementation,” p. 18.

⁵¹ 5 U.S.C. §552(a)(3).

⁵² 5 U.S.C. §552(b).

⁵³ For more discussion of the FOIA exemptions, see CRS Report R46238, *The Freedom of Information Act (FOIA): A Legal Overview*, by Benjamin M. Barczewski.

⁵⁴ 5 U.S.C. §552(f)(2)(A).

⁵⁵ 5 U.S.C. §552(a)(3)(B), P.L. 104-231, 110 Stat. 3050.

⁵⁶ DOJ, “Proactive Disclosure of Non-Exempt Agency Information.”

Under the OPEN Government Data Act, certain data assets are made known or available through the federal data catalogue without requiring a secondary step of requesting information. M-25-05 provides that certain other datasets may be made available upon request. M-25-05 explains that while many public data assets will be available through the federal data catalogue, “they may not be available there immediately or on a timeline that meets the needs of the requester” and that members of the public can continue to request access to agency data assets through FOIA. Public data assets may still be provided upon request to researchers or with adequate disclosure limitation techniques applied.⁵⁷ In such cases, the law requires agencies to provide certain metadata alongside the data asset describing “the method by which the public may access or request access to the data asset.”⁵⁸

The federal data catalogue also provides information about how researchers and other agencies may request access to certain data assets that do not qualify as public data assets and are not appropriate for dissemination. For example, under Title III of the Evidence Act, recognized statistical agencies and units (RSAUs) accept applications to access confidential statistical data for evidence-building purposes through the SAP.⁵⁹

Users seeking access to confidential data are able to search the SAP data catalogue to identify data assets that may be available for the user’s research, pending SAP approval. The metadata elements of the SAP data catalogue and the federal data catalogue are fully interoperable. Agencies with data assets included in the SAP need only create and maintain the CDI required under M-23-04 to satisfy the requirements of both Titles II and III of the Evidence Act. The federal data catalogue is the source of metadata for the SAP.

Requesting Statistical Data via the Standard Application Process⁶⁰

Federal Statistical Research Data Centers (FSRDCs) were established by the ICSP to provide researchers with access to certain restricted-use statistics produced by RSAUs.⁶¹ The FSRDC program is intended to promote coordination, quality, utility, transparency, and openness in federal statistics through increasing access to microlevel statistical data and supporting linkage of this data across different agencies. The Census Bureau, in partnership with the principal statistical agencies or units and research institutions, manages over 35 FSRDC locations across the continental United States that external researchers may request access to through the SAP.⁶² In a similar way to other request-based methods of disclosure, researchers can use federal data for certain purposes and approved projects.

The FSRDC program is governed by the FSRDC Executive Committee, which was established in July 2017. The executive committee’s stated purpose is that it

⁵⁷ OMB, “Establishment of Standard Application Process Requirements on Recognized Statistical Agencies and Units,” M-23-04, <https://www.whitehouse.gov/wp-content/uploads/2022/12/M-23-04.pdf>. See also National Science and Technology Council, *National Strategy to Advance Privacy-Preserving Data Sharing and Analytics*, Washington, D.C., March 2023, p. 19, <https://bidenwhitehouse.archives.gov/wp-content/uploads/2023/03/National-Strategy-to-Advance-Privacy-Preserving-Data-Sharing-and-Analytics.pdf>.

⁵⁸ 44 U.S.C. §3511.

⁵⁹ OMB, “Establishment of Standard Application Process Requirements.”

⁶⁰ This section was authored by Taylor R. Knoedl, Analyst in American National Government.

⁶¹ Census Bureau, “Restricted-Use Data,” <https://www.census.gov/topics/research/guidance/restricted-use-microdata.html>.

⁶² Census Bureau, “Federal Statistical Research Data Centers,” <https://www.census.gov/about/adrm/fsrdc.html>.

provides strategic vision and guidance; makes policy decisions that resolve interagency issues, capitalize on new opportunities, and strive for consensus; [and] guides transformation and provides executive sponsorship of the FSRDC program.⁶³

The executive committee charter states that membership should include the chief statistician (or his or her designee) and an institutional partner as cochairs, heads of participating federal agencies or their designees, the chief information officer of the Census Bureau, the FSRDC program director, and four representatives of the institutional partners that host the FSRDCs.⁶⁴

A researcher who wishes to access certain restricted-use data available at FSRDCs must complete an SAP then go on site to one of the FSRDC locations located at partner institutions (e.g., universities, nonprofit organizations, and Federal Reserve banks). Researchers obtaining access to restricted-use data through SAPs are required to adhere to confidentiality standards established in Title 44 of the *U.S. Code*.⁶⁵ Researchers are furthermore required to obtain special sworn status as laid out in Section 23 of Title 13 of the *U.S. Code* and are responsible for adhering to confidentiality standards established in Section 9 of Title 13.

The SAP is developed in compliance with the Confidential Information Protection and Statistical Efficiency Act of 2018 and OMB Memorandum M-23-04.⁶⁶ The guidance established in M-23-04 states that restricted-use data accessed through FSRDCs may be used only for statistical purposes where individuals are not identified.⁶⁷ Furthermore, OMB establishes standardized review criteria for project proposals:

1. Project proposals must have a statistical purpose.
2. Data use described in a project proposal must be consistent with any restrictions on use of a specific data asset.
3. Project proposals must apply sound statistical disclosure limitation models that minimize the risk of re-identification.
4. Project proposals must demonstrate that goals and objectives require use of confidential data assets.
5. Project proposals must be feasible to justify the disclosure risk of using confidential data assets.
6. Statistical agencies must assess whether project proposals support their ability to maintain trust and credibility among data users.⁶⁸

⁶³ Census Bureau, “FSRDC: Governance,” <https://www.census.gov/about/adrm/fsrdc/about/governance.html>; Census Bureau, “Restricted-Use Data Application Process,” <https://www.census.gov/topics/research/guidance/restricted-use-microdata/standard-application-process.html>.

⁶⁴ Census Bureau, “Federal Statistical Research Data Centers (FSRDC) Executive Committee Charter,” https://www.statpolicy.gov/assets/files/FSRDC%20Executive%20Committee%20Charter_%20June%202024.pdf.

⁶⁵ 44 U.S.C. §3582.

⁶⁶ 44 U.S.C. §3583(a). See OMB, “Establishment of Standard Application Process Requirements on Recognized Statistical Agencies and Units.”

⁶⁷ 44 U.S.C. §3531(12). *Statistical purposes* is defined as “the description, estimation, or analysis of the characteristics of groups, without identifying the individuals or organizations that comprise such groups” and includes “the development, implementation, or maintenance of methods, technical or administrative procedures, or information resources that support the purposes described.”

⁶⁸ OMB, “Establishment of Standard Application Process Requirements on Recognized Statistical Agencies and Units,” pp. 15-17.

Publicly available sources do not make clear whether all datasets are available at each of the different FSDRC locations.⁶⁹ The SAP can take an expected 12-24 weeks to receive a final determination, depending on whether the applicant needs data from multiple agencies or units. This process can take longer than the expected 24 weeks for applications that require approval from organizations outside of the SAP.⁷⁰

Preserving and Disposing of Data

In addition to considering whether and how information can be made available, data preservation encompasses the concept of how long and in what media or location the data will be produced. Archivists accept that some information loss of a preserved object is inevitable, resulting in an “archival sliver” of the original object. Because of this inevitability, archivists understand that the archival sliver phenomenon may distort future research.⁷¹ This section discusses records retention and disposition requirements under statute, what types of information and metadata are required to accompany data from active use within an agency to longer-term storage in NARA, considerations for when information should be removed for obsolescence or corrections, and options to provide a process for such removals.

The Federal Records Act and Records Preservation

Information must be maintained in order to be retrieved, used, and made available. While transparency statutes such as FOIA govern the availability of government information, records management statutes and programs govern the retention and preservation of certain government materials. In particular, the Federal Records Act (FRA; 44 U.S.C. Chapters 21, 29, 31, and 33), enacted in 1950 and amended since, governs the collection, retention, and preservation of federal agency records. Congress deemed *federal records* worthy of preservation because they provide information on the transaction of public business, such as the “organization, functions, policies, decisions, procedures, and essential transactions” of the government.⁷²

The FRA provides a definition of *federal records* in order to determine whether particular recorded information should be retained and managed. Whether materials meet the definition of *federal record* is based on the content of the information and not the format on which the information is stored. The definition also excludes library and museum materials made for reference or exhibition purposes and duplicate copies of records preserved only for convenience, sometimes referred to as “convenience copies.” In cases where there is disagreement over whether particular recorded information constitutes a federal record, statute expressly empowers the Archivist, the head of NARA, to determine “whether recorded information, regardless of

⁶⁹ Inter-university Consortium for Political and Social Research at the University of Michigan, “ResearchDataGov.org: About,” 2022, <https://www.researchdatagov.org/about>.

⁷⁰ Inter-university Consortium for Political and Social Research at the University of Michigan, “Frequently Asked Questions: Agency Review of Applications,” 2022, <https://www.researchdatagov.org/about>.

⁷¹ For a case study on this phenomenon, see Verne Harris, “The Archival Sliver: A Perspective on the Construction of Social Memory in Archives and the Transition from Apartheid to Democracy,” in *Refiguring the Archive*, ed. C. Hamilton et al. (Springer, 2002), pp. 135-160, https://link.springer.com/content/pdf/10.1007/978-94-010-0570-8_9.pdf. See also James A. Jacobs and James R. Jacobs, *Preserving Government Information: Past, Present, and Future* (FreeGovInfo Press, 2025), <https://freegovinfo.info/pgi>.

⁷² 44 U.S.C. §3301.

whether it exists in physical, digital, or electronic form, is a record” for purposes of the FRA and states that this determination “shall be binding on all Federal agencies.”⁷³

What Is a Federal Record?

44 U.S. Code §3301

Records include “all recorded information, regardless of form or characteristics, made or received by a Federal agency under Federal law or in connection with the transaction of public business and preserved or appropriate for preservation by that agency or its legitimate successor as evidence of the organization, functions, policies, decisions, procedures, operations, or other activities of the United States Government or because of the informational value of data in them.”

Records **do not include** “library and museum material made or acquired and preserved solely for reference or exhibition purposes” or “duplicate copies of records preserved only for convenience.”

Maintaining and Disposing of Federal Records

The FRA also governs the length of time for which federal records are to be maintained and if or when federal records materials may be destroyed. Materials are assessed for their preservation value through the records control schedule process.⁷⁴ A records schedule is created by agencies in consultation with NARA and provides a disposition authority for the set of records discussed in the schedule. The disposition authority provides information on where the information should be stored and if and when the information should be destroyed.

A records schedule can be any of the following:

- A standardized form (SF 115) that NARA has approved to authorize the disposition of federal records (i.e., disposition authority)⁷⁵
- A General Records Schedule (GRS) issued by NARA, which authorizes, after specified periods of time, the destruction of temporary records or the transfer of permanent records to the Archives that are common to several or all agencies⁷⁶
- A published agency manual or directive containing the records descriptions and disposition instructions approved by NARA on one or more standardized forms or issued by NARA in the GRS

All federal records must be covered by a NARA-approved records schedule or a GRS. The records schedule should include a description of each type or series of records and note whether the records are temporary (to be discarded by the federal government) or permanent (to be permanently retained by NARA). Generally, temporary records (approximately 95%-98% of federal records) need to be maintained for only a limited period of time and stay within the custody of the creating agency, while each permanent record (2%-5% of federal records) is accessioned for continuing preservation at NARA at a point in time determined by the records

⁷³ 44 U.S.C. §3301(b). Prior to the adoption of the Presidential and Federal Records Act Amendments of 2014 (P.L. 113-187), the statutory definition of *federal record* included certain types of materials or platforms on which records could be created or captured, such as “books, papers, photographs,” and “machine-readable formats.” According to the accompanying Senate report, the act amended the definition of *federal record* to include the phrase *regardless of form or characteristics* in order to “shift the emphasis away from the physical media used to store information to the actual information being stored.” See also CRS In Focus IF11119, *Federal Records: Types and Treatments*, by Meghan M. Stuessy.

⁷⁴ NARA, “Records Control Schedules (RCS),” <https://www.archives.gov/records-mgmt/rcs>.

⁷⁵ A copy of SF 115 may be located at <https://www.gsa.gov/system/files/2025-03/SF115-91.pdf>.

⁷⁶ See also NARA, “What Are the General Records Schedules (GRS),” <https://www.archives.gov/records-mgmt/grs>.

control schedule.⁷⁷ For a permanent record, the schedule includes the date the record is to be transferred to NARA. Unless and until NARA accessions records, records maintenance responsibilities lie with the creating agency. Subsequent requirements for an agency to supply preservation metadata to NARA are observable only at the time the record is accessioned and would otherwise rely on a strong records management program within the agency.

Records schedules must be cleared by internal agency stakeholders, the Government Accountability Office (when required by Title 43, Section 1225.20(a), of the *Code of Federal Regulations*), and NARA. Disposition instructions approved by NARA are mandatory.⁷⁸ In addition, NARA must publish a notice of agency requests for the disposal of records in the *Federal Register*.⁷⁹ If NARA has previously approved a records schedule to dispose of certain agency records, a notice is published only if the proposed retention period is shorter. The publication of these notices allows interested persons to submit written comments on the records to NARA before disposal is approved or reapproved with a shorter retention period.

Technical Details Required to Archive Data

Recalling the earlier discussion of the information life cycle, one may better understand requirements for agency data by determining which stage the data is in.⁸⁰ The third stage concerns storing and disposing of information and has implications for continued information availability over time, but it is connected to and influenced by the two earlier stages concerning how data is (1) collected and created and (2) used or shared.

In particular, OMB Circular A-130 requires that “[r]ecords management functions and retention and disposition requirements must be fully incorporated into information life cycle processes and stages, including the design, development, implementation, and decommissioning of information systems, particularly Internet resources to include storage solutions and cloud-based services.”⁸¹ Records management practices implicate the ability of agencies to preserve and make data available, and NARA notes that, with respect to structured data, records management “should be considered in all aspects of the database lifecycle.”⁸² Agencies are also required to maintain up-to-date documentation about their electronic information systems, including:

- all technical characteristics necessary for reading and processing the records contained in the system;
- restrictions on access and use of the information; and
- update cycles or conditions and rules for adding, changing, or deleting information in the system.⁸³

If agency data records warrant permanent preservation, NARA provides certain criteria that are required to accompany the agency transfer of electronic records to the archives. In particular, agencies “must transfer documentation adequate to identify, service, and interpret the permanent

⁷⁷ NARA, “National Records Management Training Program,” February 2021, <https://www.archives.gov/files/r1-006-tipsheets-permanenttemporarynonrecord-1.pdf>.

⁷⁸ 44 U.S.C. §3314.

⁷⁹ 44 U.S.C. §3303a(a).

⁸⁰ OMB, *Managing Information as a Strategic Resource*, Circular A-130, p. 29.

⁸¹ 44 U.S.C. §2901(2) and OMB, *Managing Information as a Strategic Resource*, Circular A-130, p. 14.

⁸² NARA, *System Inspection (Multi-Agency Report): Structured Data Managed Within Databases*, November 2024, p. 6, <https://www.archives.gov/files/records-mgmt/resources/structured-data-managed-within-databases-system-mai.pdf>.

⁸³ 36 C.F.R. §1236.26.

electronic records,” and documentation for data files and data bases “must include record layouts, data element definitions, and code translation tables (codebooks) for coded data. Data element definitions, codes used to represent data values, and interpretations of these codes must match the actual format and codes as transferred.”⁸⁴ Further, “Data files and databases must be transferred to the National Archives of the United States as flat files or as rectangular tables; *i.e.*, as two-dimensional arrays, lists, or tables.”⁸⁵

The information required to accompany data files and databases often takes the form of *metadata*. Statute broadly defines *metadata* as “structural or descriptive information about data such as content, format, source, rights, accuracy, provenance, frequency, periodicity, granularity, publisher or responsible party, contact information, method of collection, and other descriptions.”⁸⁶ With respect to the transfer of records, NARA requires specific *transfer metadata*, including access restrictions, the name of the creating organization, and use restrictions, among other information.⁸⁷

To enable continued digital preservation of electronic records, NARA’s preservation action plan for structured data and databases requires that the information be accompanied with a “documented data model.”⁸⁸ NARA explains that the data model describes “what categories of data will be stored in which fields, columns or tags; data types (numeric, currency, alphabetic, name, date, address); and, if possible, controlled vocabulary” and, importantly, that the documentation of the structure “enables retention of the database behavior” to support the data’s continued usability.⁸⁹ After NARA has accessioned records from agencies, records may be requested and provided subject to FOIA exemptions and NARA restrictions.⁹⁰

Removing Data

Under Section 3506 of Title 44 of the *U.S. Code*, an agency head is required to “improve the integrity, quality, and utility of information to all users within and outside the agency” as part of his or her information resources management responsibilities. Arguably, this could include agency heads removing or deprioritizing outdated or inaccurate information from public dissemination. This may also be construed as part of an agency’s responsibilities to efficiently manage information resources by emphasizing current data over historical data. However, such removal could conflict with other statutory responsibilities to maintain records demonstrating government policy and decisionmaking processes.

While the statutory language of the OPEN Government Data Act discusses when and how information is to be added to the CDI and federal data catalogue, the statute does not provide information on whether and how information may be removed from such resources. Agency heads are required to provide descriptions of data assets, including the date on which the data assets were most recently updated, and further requires agencies to update their data inventories to include additional data assets “not later than 90 days after the date of such creation or

⁸⁴ 36 C.F.R. §1235.48.

⁸⁵ 36 C.F.R. §1235.50(b).

⁸⁶ 44 U.S.C. §3502(19).

⁸⁷ 36 C.F.R. §1236.58.

⁸⁸ NARA, “U.S. National Archives and Records Administration Digital Preservation Framework,” June 27, 2025, <https://github.com/usnationalarchives/digital-preservation>.

⁸⁹ NARA, “Preservation Action Plan for Structured Data/Databases,” November 18, 2024, https://github.com/usnationalarchives/digital-preservation/blob/master/Digital_Preservation_Record_Categories/NARA_PreservationActionPlan_Calendars.md.

⁹⁰ 36 C.F.R. §1256.

identification.”⁹¹ Although researchers and the public may derive value from being able to trace data over time to determine changes in trends or collection methods, the statute does not explicitly consider versioning requirements for data assets. However, requiring these attributes for Data.gov may help address or clarify difficulties in “Measuring Data Availability,” as discussed below.

Similarly, Memorandum M-25-05 does not discuss versioning requirements where the public could view multiple iterations of a data asset over time at a static location. However, the memorandum does suggest that agencies “should establish processes to remove references to data assets that have been [disposed of] pursuant to applicable records retention and disposition policies,” although such processes would not necessarily be uniform across agencies.⁹² OMB does remind agencies that they are to engage the public and encourage collaboration by “providing the public with the opportunity to request specific data assets to be prioritized for disclosure and to provide suggestions for the development of agency criteria with respect to prioritizing data assets for disclosure.”⁹³ Again, this text stresses agency responsibilities for disclosure, not data removal.

Corrections Under the Information Quality Act

Enacted in 2001, the Information Quality Act (IQA) directs the OMB director to issue government-wide guidelines that “provide policy and procedural guidance to Federal agencies for ensuring and maximizing the quality, objectivity, utility, and integrity of information (including statistical information) disseminated by Federal agencies.”⁹⁴ In OMB’s final guidelines from 2002, OMB defines *quality* as comprising information utility, objectivity, and integrity.⁹⁵ With respect to *objectivity*, information disseminated by agencies is to be presented in an “accurate, clear, complete, and unbiased manner,” and *integrity* refers to information security, or “information from unauthorized access or revision, to ensure that the information is not compromised through corruption or falsification.”⁹⁶

The final guidelines require agencies to establish administrative mechanisms allowing “affected persons to seek and obtain correction of information maintained and disseminated by the agency” that does not comply with OMB guidelines.⁹⁷ The guidelines suggest that agencies should design these administrative mechanisms to “facilitate public review,” although implementation may vary among agencies.

OMB explains, “The primary benefit of public transparency is not necessarily that errors in analytic results will be detected, although error correction is clearly valuable. The more important benefit of transparency is that the public will be able to assess how much an agency’s analytic result hinges on the specific analytic choices made by the agency.”⁹⁸ The guidelines require that agencies annually report on the number, nature, and resolution of complaints received, although

⁹¹ 44 U.S.C. §3511(a)(2)(A) and (a)(3).

⁹² OMB, “Phase 2 Implementation,” p 11.

⁹³ 44 U.S.C. §3506(d)(6)(B).

⁹⁴ P.L. 106-554; 114 Stat. 2763A–154.

⁹⁵ OMB, “Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies; Republication,” 67 *Federal Register* 8453, February 22, 2002. *Utility* refers to the usefulness of the information to the intended users.

⁹⁶ 67 *Federal Register* 8452.

⁹⁷ 67 *Federal Register* 8452.

⁹⁸ 67 *Federal Register* 8456.

information may be corrected without a corresponding notification until the next reporting cycle, leading to a lack of transparency surrounding the information's current quality.⁹⁹

In 2019, OMB issued Memorandum M-19-15 suggesting it would take a more active role in requests for correction under the IQA. In it, OMB requires agencies to revise their procedures so as to not take more than 120 days to respond to a correction and additionally that agencies should share draft responses with OMB prior to releasing corrections.¹⁰⁰ While a decentralized correction process may make it difficult to identify individual information corrections, centralizing the process through OMB may grant OMB additional influence in whether and how information should be altered. Additionally, neither the statute nor the final guidelines nor corresponding OMB memoranda detail information removal as an outcome of the correction process.

Removal of Convenience Copies and Web Records

As previously discussed, the FRA makes a distinction between recordkeeping copies of information, which must be maintained and preserved, and convenience copies, which are duplicate copies. In the case of copies of electronic records that may be accessible on an agency's website (such as documents, spreadsheets, or audio/visual files), these copies may be removed from the website without contravening the FRA so long as the recordkeeping copy remains and is maintained in accordance with required records management practices.

NARA differentiates between information supplied on agency websites and the web pages themselves.¹⁰¹ While recordkeeping files of information on agency websites may be governed through content-specific agency records schedules, web pages themselves may be inconsistently managed. In 2005, NARA explained that there are "currently no items in the [GRSs] that were developed to specifically cover web records" and that agencies may rely on a mixture of GRSs to manage web records. An opt-in GRS is available government-wide for agencies to simplify their records management processes. In the event a GRS does not cover or is insufficient for a particular agency's information type, agencies must work with NARA to create particularized records control schedules for the information.

Because a single GRS does not apply to agency web records, agencies must either blend together existing GRS guidance or create specific website schedules for web records. In corresponding 2005 documentation, NARA explains that "For the sake of simplicity and ease of management of the web site, an agency may also choose to use a single item and retention period for web records even if there are variations in business needs and risk."¹⁰² Because of the proliferation of website use by agencies, it may be difficult for agencies to balance the usability of websites and provide up-to-date information against providing a consistent user experience for web items. NARA

⁹⁹ In a 2015 report, GAO recommended that OMB issue guidance providing "specific time frames for agencies to post information on the IQA correction requests they had received," and as of 2026, GAO marks this recommendation as still open. See GAO, *Information Quality Act: Actions Needed to Improve Transparency and Reporting of Correction Requests*, GAO-16-110, December 21, 2015, p. 34, <https://www.gao.gov/products/gao-16-110>. In OMB Memorandum M-19-15, OMB observes that agencies frequently unilaterally extend their own deadlines for replying to requests for correction, "taking a year or more to provide a substantive response." OMB prescribes that agencies "should set and adhere to reasonable timelines, not to exceed 120 days, for a response without the concurrence of the requester" but does not provide a separate enforcement mechanism. See OMB, "Improving Implementation of the Information Quality Act," M-19-15, April 24, 2019, p. 10, <https://www.whitehouse.gov/wp-content/uploads/2019/04/M-19-15.pdf>.

¹⁰⁰ OMB, "Improving Implementation of the Information Quality Act."

¹⁰¹ NARA, "Frequently Asked Questions about GRS 5.1, Common Office Records," April 2024, <https://www.archives.gov/records-mgmt/grs/faqs-for-grs-5-1>.

¹⁰² NARA, "NARA Guidance on Managing Web Records," January 2005, p. 22, <https://www.archives.gov/files/records-mgmt/pdf/managing-web-records-index.pdf>.

suggests that although agencies do not necessarily need to keep all web pages and changes for long periods of time, agencies should manage website content to mitigate risk to business operations by capturing their websites on a periodic basis with accompanying site maps.¹⁰³ As the amount of agency information in electronic formats continues to grow, agencies may also feel pressure to prioritize maintenance of electronic recordkeeping copies of files over website interfaces and configuration changes.

Although these procedures and documentation requirements apply to both temporary and permanent records created by an agency, NARA has provided specific file format and metadata requirements for permanent records to be transferred to NARA for continuing preservation. Many of these preferred and required file formats comport with Web ARChive, International Organization for Standardization, and International Electrotechnical Commission standards.¹⁰⁴ Agencies may prioritize implementing these standards only for records deemed to become permanent records.¹⁰⁵

Measuring Data Availability

Consumers of federal data have myriad ways to locate and use relevant datasets through voluminous agency and programmatic websites, via information request, or through on-site visits. Simultaneously, the landscape of data access, consistent versioning, and formatting is constantly changing. In light of the earlier discussion of the components of data availability—specifically as it relates to timely and reliable access to government data—Congress, the public, and stakeholders may question whether there are authoritative ways to measure how data availability may have changed over time. Congress may also be interested in determining whether there are trends to certain data becoming available or when they are altered and removed. Such trends may provide insight and direction for Congress to further inspect agency activities or make decisions to support new data use cases. However, as noted above, the lack of requirements to include versioning information—such as when or how data has been modified and access to previous versions of the information to provide a comparison to other copies for agency data—can create difficulties in efforts to measure the availability of federal data over time.¹⁰⁶

For example, Data.gov, which is maintained by the General Services Administration, functions as a data catalogue where users can search for and locate sources of government data.¹⁰⁷ The agency explains that “Data.gov does not host data directly (with a few exceptions), but rather aggregates metadata about open data assets in one centralized location. Once an agency creates an open data source with the necessary format and metadata requirements, the Data.gov team can harvest the metadata directly, synchronizing that source’s metadata on Data.gov as often as every 24 hours.”¹⁰⁸ Because of the decentralized nature of federal data hosting across various websites, the Data.gov topline number may be one way to assess the changing availability of federal data

¹⁰³ NARA, “NARA Guidance on Managing Web Records,” pp. 6 and 14.

¹⁰⁴ NARA, “Appendix A: Tables of File Formats,” August 2025, <https://www.archives.gov/records-mgmt/policy/transfer-guidance-tables.html>. See also International Internet Preservation Consortium, “The WARC Format 1.0,” <https://iipc.github.io/warc-specifications/specifications/warc-format/warc-1.0/index.html>.

¹⁰⁵ NARA, “Universal Electronic Records Management (ERM) Requirements,” June 23, 2023, <https://www.archives.gov/records-mgmt/policy/universalermsrequirements>.

¹⁰⁶ See “Removing Data” above.

¹⁰⁷ 44 U.S.C. §3511. For more discussion of the development of Data.gov, see the “The OPEN Government Data Act” section above.

¹⁰⁸ GSA, “How to Get Your Open Data on Data.gov,” <https://resources.data.gov/resources/data-gov-open-data-howto/>.

government-wide.¹⁰⁹ However, it is equally possible that because information is stored locally by agencies, it may not reliably be captured or reported by Data.gov. In some cases, only users who frequently look for the data on agency websites, rather than Data.gov's catalogue, may notice whether the data is missing.

Media reports suggested that, between January 21 and January 30, 2025, data was being removed from Data.gov and agency websites, with one source reporting that Data.gov's topline number of available datasets reduced by 2,000 (out of more than approximately 305,000).¹¹⁰ Some observers are also tracking the removal of specific datasets, variables, and tools.¹¹¹ A possible consequence of these reported changes in data availability is reduced confidence in and transparency into federal government operations: The National Press Club wrote, "Public data belongs to the public. When that information is removed, restricted, or allowed to quietly vanish, it undermines transparency and weakens accountability. It also makes it harder for those in vital industries including public health, education, law enforcement, transportation, and journalism to serve the public."¹¹²

CRS examined the availability of federal data reported to Data.gov (not individual agency websites) using the Internet Archive's Wayback Machine, a publicly accessible archive of website content changes chronicled by the Internet Archive's web crawling software.¹¹³ A *web crawler* or *web harvester* is a type of software that automatically and methodically downloads, indexes, and stores content from the web.¹¹⁴

To identify this data, CRS examined Data.gov "metrics" pages that were downloaded through web crawls by the Wayback Machine. These archived metrics pages contain various measurements (i.e., raw datasets, geospatial datasets, and datasets) of federal agency participation on Data.gov over time. In reviewing these web crawls of Data.gov at different points in time since its first capture on April 4, 2009, the amount and level of detail in reporting on available

¹⁰⁹ Other organizations preserve the interface of government websites on a piecemeal basis, potentially allowing another method of determining how data hosting has changed. Libraries and research organizations have worked together to preserve material from U.S. government websites during presidential transitions, although the scope of federal agency involvement has varied. For example, NARA has focused on capturing and archiving congressional websites and presidential websites. The Library of Congress has focused on thematic collections, including legislative branch agencies and U.S. national election campaigns. See Caralee Adams, "Update on the 2024/2025 End of Term Web Archive," The Internet Archive, February 6, 2025, <https://blog.archive.org/2025/02/06/update-on-the-2024-2025-end-of-term-web-archive/>.

¹¹⁰ Jason Koebler, "Archivists Work to Identify and Save the Thousands of Datasets Disappearing from Data.gov," *404 Media*, January 30, 2025, <https://www.404media.co/archivists-work-to-identify-and-save-the-thousands-of-datasets-disappearing-from-data-gov/>; Rachel Santarsiero, "Disappearing Data: Trump Administration Removing Climate Information from Government Websites," National Security Archive, February 6, 2025, <https://nsarchive.gwu.edu/briefing-book/climate-change-transparency-project-foia/2025-02-06/disappearing-data-trump>; Katherine J. Wu, "CDC Data Are Disappearing," *The Atlantic*, January 31, 2025, <https://www.theatlantic.com/health/archive/2025/01/cdc-dei-scientific-data/681531/>; Will Stone and Selena Simmons-Duffin, "Trump Administration Purges Websites Across Federal Health Agencies," *National Public Radio*, January 31, 2025, <https://www.npr.org/sections/shots-health-news/2025/01/31/nx-s1-5282274/trump-administration-purges-health-websites>.

¹¹¹ America's Essential Data, "Confirmed Data Terminations and Removals," December 1, 2025, <https://essentialdata.us/in-memoriam.html>; America's Data Index, "Data Checkup: A Health Check for Federal Data Collections," January 31, 2026, <https://dataindex.us/collections/>.

¹¹² National Press Club, "National Press Club Statement on the Elimination of Data from Federal Agency Websites," press release, March 16, 2026, <https://www.press.org/newsroom/national-press-club-statement-elimination-data-federal-agency-websites>.

¹¹³ The Internet Archive's Wayback Machine is located at <https://web.archive.org/>.

¹¹⁴ Digital Preservation Coalition, "Digital Preservation Handbook: Glossary," 2015, <https://www.dpconline.org/handbook/glossary#C>. GSA provides documentation for the Data.gov web harvester and harvesting logic at GSA, "GSA/datagov-harvester," <https://github.com/GSA/datagov-harvester>.

data has fluctuated. A capture from January 3, 2012, indicates that 390,177 raw and geospatial datasets are available, compared to 378,529 on January 1, 2013.¹¹⁵ As of December 2025, Data.gov reports having approximately 364,410 datasets available.¹¹⁶

CRS also attempted to review agency and sub-agency dataset and visitor numbers intermittently published by Data.gov on their metrics pages since January 15, 2014. However, the underlying agency and sub-agency datasets and visitor numbers were not available through the Wayback Machine before 2014.¹¹⁷ Although these numbers have fluctuated since Data.gov's inception, due to limited visibility into how datasets are being counted and how frequently such numbers and associated metadata are being updated, it is difficult to conclusively determine why such fluctuations occur.

In its current iteration, the Data.gov metrics page appears to have been populated since August 1, 2024.¹¹⁸ Monthly reports on datasets available by organization (e.g., federal agency, state, and local governments) are available to download from that date through July 2025 by using the links provided on individual Wayback Machine web crawls. There does not currently appear to be publicly available, compiled information at the organization level aside from this method. In reviewing the monthly reports, CRS determined that the number of datasets at the organization level varied over time: Some organizations steadily added to their counts, while others appear to have changed a handful of datasets. Some organizations may contribute a dozen datasets or fewer; other organizations contribute tens of thousands of datasets.

Notably, datasets from six federal organizations appear to have been removed entirely from the Data.gov monthly reports as of February 4, 2026: the U.S. Geological Survey,¹¹⁹ the Office of Navajo and Hopi Indian Relocation,¹²⁰ the Federal Emergency Management Agency,¹²¹ the U.S. Agency for International Development,¹²² the U.S. International Trade Commission,¹²³ and the Department of Veterans Affairs.¹²⁴ Data for the Government Printing Office and the National Endowment for the Arts appeared to be removed in certain reports but reappear in the latest January 31, 2026, report.¹²⁵ While the overall data numbers are difficult to replicate, the absence

¹¹⁵ GSA, "Data.gov," archived January 3, 2012, <https://web.archive.org/web/20120103115422/http://www.data.gov/> and GSA, "Data.gov," archived January 1, 2013, <https://web.archive.org/web/20130101010506/http://www.data.gov/>. The terms *raw datasets* and *geospatial datasets* may contribute to a change in counting datasets seen on later iterations of Data.gov.

¹¹⁶ GSA, "Data.gov," archived December 9, 2025, <https://web.archive.org/web/20251209151746/https://data.gov/>.

¹¹⁷ GSA, "Federal Agency Participation—Data.gov," archived March 31, 2022, <https://web.archive.org/web/20220331215801/https://data.gov/metrics/>; GSA, "Data.gov Visitor Metrics," archived July 13, 2022, <https://web.archive.org/web/20220713163400/https://catalog.data.gov/dataset/data-gov-visitor-metrics>.

¹¹⁸ GSA, "Data.gov Metrics Dashboard," archived August 1, 2024, <https://web.archive.org/web/20240801043010/https://data.gov/metrics/>. Viewing prior crawls from July 2024 appears to redirect the site to <http://www.data.gov/dashboard>, which is now inactive, rather than <http://data.gov/metrics>.

¹¹⁹ The U.S. Geological Survey data is missing from the February 28, 2025, report; reappears for the October 31, 2025, report; and disappears again in the November 30, 2025, report.

¹²⁰ The Office of Navajo and Hopi Indian Relocation data is missing from the April 20, 2025, report; reappears for the October 31, 2025, report; and disappears again in the November 30, 2025, report.

¹²¹ The Federal Emergency Management Agency data is missing from the July 31, 2025, report; reappears in the October 31, 2025, report; and disappears again in the November 30, 2025, report.

¹²² The U.S. Agency for International Development data is missing from the September 30, 2025, report. It reappears in the October 31, 2025, report but disappears again in the November 30, 2025, report.

¹²³ The U.S. International Trade Commission data is missing from the January 31, 2026, report.

¹²⁴ The Department of Veterans Affairs data is missing from the January 31, 2026, report.

¹²⁵ Both the Government Publishing Office and National Endowment for the Arts appear for the first time in the (continued...)

of certain organizations that were previously included indicates that some data are no longer available on Data.gov. However, it does not indicate why the data are no longer available through the site or whether the organizations that previously provided the data are continuing to maintain the data in other locations.

Although Data.gov provides individual landing pages for datasets, it does not provide archived versions of the data, relying instead on agency records management practices to provide such information access. Because Data.gov functions as a catalogue that directs users to agency-hosted data, the data's catalogue entry could also remain while the dataset itself could be removed by the agency. Overall, individual public dataset loss may depend on being noticed by data users reviewing individual datasets rather than current reporting mechanisms.

The corresponding House Committee on Oversight and Government Reform report supporting the passage of the OPEN Government Data Act describes the legislation as creating a comprehensive inventory of federal data assets that “would allow researchers to quickly ascertain the scope of data products available,” although in practice the information provided on Data.gov is likely incomplete.¹²⁶ This discrepancy between the stated intent from the House report of creating a comprehensive accounting of data and practical difficulties in implementing such a tool to achieve this goal—whether Data.gov or another tool—is substantial and likely to persist.

When considering Data.gov or other data inventories, Congress may wish to examine whether the current reporting mechanisms used by Data.gov allow stakeholders to evaluate whether it sufficiently meets the goals provided in the OPEN Government Data Act and consideration of changes in statute or implementation of Data.gov. Although at present the statute requires agencies to inventory their data, the statute does not specify versioning or monthly reporting requirements for the site that may better enable data availability tracking.

The significance of data availability numbers may also change when viewed within the context of their limitations, as demonstrated through current Data.gov metrics. For example, the existing monthly reports on Data.gov do not appear to provide sufficient information to determine how the datasets it collects change over time. It is possible, for example, that datasets could be taken down from view and replaced, unaltered, at another date. It is also possible, given the information available in the reports, that a certain number of datasets could be taken down and replaced with a different but equal number of datasets without changing the numbers in the reports. As discussed later in this report, Congress may consider adopting certain data versioning, fixity, and governance procedures to mitigate against such misinterpretations.

Independent Efforts to Archive Data

In response to reports that federal data and statistics were being removed from online public availability, private entities created independent efforts to preserve government information.¹²⁷ For example, the Data Rescue Project describes itself as a clearinghouse for preserving at-risk public information: “What began as a simple Google Sheet has evolved into a crowdsourced tool that catalogues and coordinates data preservation efforts across institutions.”¹²⁸ Another, created

Data.gov reports in September 30, 2025; disappear for October 31, 2025; and reappear in the November 30, 2025, report.

¹²⁶ U.S. Congress, House Committee on Oversight and Government Reform, *Foundations for Evidence-Based Policymaking Act of 2017*, committee print, 115th Cong., 1st sess., November 15, 2017, H.Rept. 115-411, p. 12, <https://www.congress.gov/115/crpt/hrpt411/CRPT-115hrpt411.pdf>.

¹²⁷ Koebler, “Archivists Work to Identify.”

¹²⁸ Data Rescue Project, “About the Mission,” <https://portal.datarescueproject.org/>.

by the Harvard Law School Library Innovation Lab, has provided a mirror of all data files linked from Data.gov since November 2024.¹²⁹

America's Data Index seeks to monitor federal data infrastructure “from dataset availability and new releases to planned and unplanned changes to collections” by providing ways to monitor changes to (or the *fixity* of) government data.¹³⁰ *Fixity* describes an assurance that the digital file has remained unaltered. Digital preservationists perform fixity checks, or ways of verifying the digital fingerprint of a digital object, to ensure that the digital object has not changed over time or when being sent to other users.¹³¹ It also monitors information collection request changes when agencies submit requests to OMB to create, renew, modify, or discontinue information collections as a potential early indicator that available data may change or become unavailable in the future.¹³²

In some ways, these independent efforts indicate adoption of the “lots of copies keep stuff safe” (or LOCKSS) digital preservation principle, where decentralized copies of information can reduce central points of software, hardware, or human failure.¹³³ As a 2024 American Statistical Association report suggested, there is value in government data supplementing private and business data, and there may also be roles for both government and private archives of federal data.¹³⁴

Issues for Congress

Throughout 2025, media reports suggested that the availability of federal data was reduced, though it is difficult using existing reporting on federal data to systematically determine how, when, or why government data might be removed or altered.¹³⁵ As discussed, some of this difficulty may be attributed to challenges related to tracing individual data assets and changes to them, preservation requirements, and dissemination and availability requirements, among others. In January 2026, Data.gov began testing a new catalogue interface and providing more

¹²⁹ Harvard Law School, Library Innovation Lab, “Announcing the Data.gov Archive,” February 6, 2025, <https://lil.law.harvard.edu/blog/2025/02/06/announcing-data-gov-archive/>. See also Raphael Satter, “Harvard Law Library Acts to Preserve Government Data amid Sweeping Purges,” Reuters, February 6, 2025, <https://www.reuters.com/world/us/harvard-law-library-acts-preserve-government-data-amid-sweeping-purges-2025-02-06/>; and Source Cooperative, “Archive of Data.gov,” December 13, 2024, <https://source.coop/repositories/harvard-lil/gov-data/description>.

¹³⁰ America's Data Index, “America's Data Index,” <https://dataindex.us>.

¹³¹ Digital Preservation Coalition, “Digital Preservation Handbook: Fixity and Checksums,” 2015, <https://www.dpconline.org/handbook/technical-solutions-and-tools/fixity-and-checksums>. See also National Digital Stewardship Alliance, “Checking Your Digital Content: What Is Fixity, and When Should I Be Checking It?,” 2014, <https://www.digitalpreservation.gov/documents/NDSA-Fixity-Guidance-Report-final100214.pdf>.

¹³² America's Data Index, “Information Collection Request (ICR) Tracker,” <https://dataindex.us/icr>. See also CRS In Focus IF11837, *The Paperwork Reduction Act and Federal Collections of Information: A Brief Overview*, by Maeve P. Carey and Natalie R. Ortiz.

¹³³ LOCKSS Program at Stanford University, “Preservation Principles,” <https://www.lockss.org/about/preservation-principles>.

¹³⁴ American Statistical Association, “The Nation's Data at Risk: Meeting America's Information Needs for the 21st Century,” July 9, 2024, p. 2, <https://www.amstat.org/docs/default-source/amstat-documents/the-nation-s-data-at-risk-supporting-materials/valueoffederalstatistics.pdf>.

¹³⁵ Koebler, “Archivists Work to Identify”; Santarsiero, “Disappearing Data”; Wu, “CDC Data Are Disappearing”; Stone and Simmons-Duffin, “Trump Administration Purges Websites.”

information on when data asset metadata were recorded, which may help users interact with the catalogue and understand how recently data assets were harvested.¹³⁶

Nonetheless, Congress may wish to reexamine existing policy ambiguities and the underlying statutory structure incentivizing agencies toward certain institutional decisions and disclosure procedures to determine whether modernization is needed. Using Data.gov as an example, Congress might conduct oversight of agency practices and consider additional specificity and implementation of data stewardship and risk management practices to guard against data loss and misuse. Generally, these practices coalesce around the concepts of web capture, data fixity and provenance, and data governance and transparency.

Web Capture

Congress might consider the role and efficacy of regularized web captures of agency websites. As demonstrated through the Data.gov example, although Data.gov serves as a catalogue of a significant quantity of available federal data, it does not record changes to datasets over time, and it does not host the underlying information from agencies that may be housed on a variety of changing agency websites. Preservation organizations have previously cooperated to preserve certain government websites on an approximately four-year basis, such as through the “End of Term” web archive projects, and smaller changes may be captured by individuals leveraging the Internet Archive’s on-demand web crawler.¹³⁷ Critics have also noted that heavy reliance on the Internet Archive may unintentionally make the service a single point of failure should any of the materials hosted there be removed or destroyed.¹³⁸ While Congress could allow this blend of web archiving practices to continue and determine that the current loss of certain content is acceptable, Congress could also require agencies to capture and preserve web content on a more frequent basis, either within individual agency records management programs or in compliance with revised NARA guidance and practice. Rather than relying on archiving entities to spearhead government-wide preservation of web content every few years, Congress could mandate annual or quarterly web content preservation by agencies as part of agency records management protocols.

To balance against needs to present current information in such a way that it is not overshadowed by obsolete or historical information, Congress could also consider that such policies would require a consistent agency web repository for older web captures and datasets. Although most government records are considered to warrant temporary preservation and are not considered permanent records to be transferred to NARA, Congress could require that datasets provided publicly via agency websites be versioned and accessible until the dataset is to be destroyed at the end of its maintenance period per the dataset’s records control schedule. This could permit agencies to continue to focus resources on current information while also allowing for obsolete information to be removed in accordance with existing procedures.

Data Fixity and Provenance

Congress could also require safeguards related to data fixity and documenting the provenance of federal data throughout the data’s life cycle. Data *provenance* refers to the documentation of a

¹³⁶ See catalog-beta.data.gov and harvest.data.gov.

¹³⁷ Caralee Adams, “Update on the 2024/2025 End of Term Web Archive,” Internet Archive, February 6, 2025, <https://blog.archive.org/2025/02/06/update-on-the-2024-2025-end-of-term-web-archive/>.

¹³⁸ Chris Stokel-Walker, “We’re Losing Our Digital History. Can the Internet Archive Save It?,” BBC, September 16, 2024, <https://www.bbc.com/future/article/20240912-the-archivists-battling-to-save-the-internet>.

dataset's origin, edit history, and ownership and access controls.¹³⁹ Data provenance is related to the concept of data quality. However, it is separate from assessments of the accuracy of the datasets' content and instead refers to the chain of custody involved in the creation, handling, and provision of the data. Ensuring the data's provenance can also improve the transparency and reliability of information produced by subsequent uses of the data, including artificial intelligence applications.

Regardless of the data's quality, Congress and the public may prefer to use information that is verified to be genuine, unaltered federal data. Archival practices currently exist to assure that digital files remain untampered with and unmodified throughout preservation processes, but these protocols may take place only on data that warrants continued preservation and not on all data currently being used by agencies or the public. Fixity instruments, such as checksums and cryptographic hashes, can be used to check that datasets have not been changed from the last time they were accessed and can be used to provide unique digital fingerprints specific to individual datasets.¹⁴⁰ A single character change in a dataset would produce a new digital fingerprint, thereby allowing users to verify the fixity and reliability of the specific dataset being used.

In the case of Data.gov, each entry provides certain metadata on the data asset, including a cryptographic source hash for the provided data asset. Although this is helpful information to users, it is of little utility at present, because the Data.gov entry does not provide a history of source hashes, meaning that users can only check that the datasets they are using match the datasets currently being provided by agencies absent a record of prior hashes.¹⁴¹

While preserving and tracking the underlying dataset is one component of ensuring the replicability of federal findings, researchers are increasingly adopting digital object identifiers (DOIs) to provide more durable and traceable access to articles and documents. As information is increasingly provided through electronic means and because websites can break, DOIs allow anyone to use the provided permanent web address to find referenced research material regardless of web address changes.¹⁴² Congress could consider mandating a combination approach for electronic agency information by focusing not only on the accuracy of individual datasets but also on the permanence of agency findings by assigning DOIs to agency reports or publications.

Data Governance and Transparency

Congress might also consider conducting oversight on whether and how agencies are following their information resources management responsibilities as dictated in statute. Components of information resources management include agencies tracking and maintaining information throughout its life cycle, predictably disseminating agency information, and ensuring that release data is accurate and objective. Although these responsibilities have existed in statute for decades, NARA's 2024 inspection of agency management of structured data managed within databases

¹³⁹ Shay HersHKovitz and Corinna Turbes, "The Imperative of Data Provenance in AI," Data Foundation, September 24, 2025, p. 3, <https://datafoundation.org/news/reports/697/697-Data-Provenance-in-AI>.

¹⁴⁰ Digital Preservation Coalition, "Digital Preservation Handbook: Fixity and Checksums." See also National Digital Stewardship Alliance, "Checking Your Digital Content."

¹⁴¹ See, for example, Department of Health and Human Services, "U.S. Chronic Disease Indicators," February 3, 2025, <https://catalog.data.gov/dataset/u-s-chronic-disease-indicators>.

¹⁴² See University of Illinois-Chicago, "What Is a DOI and How Do I Use Them in Citations?," <https://ask.library.uic.edu/faq/345899>; Chicago Manual of Style, "14.8: Digital Object Identifiers (DOIs)," <https://www.chicagomanualofstyle.org/book/ed17/part3/ch14/psec008.html>. "A URL based on a DOI, which will always direct readers to information about the source, if not full access to it, should be preferred where available" (Chicago Manual of Style, "14.11: Library and Other Bibliographic Databases," <https://www.chicagomanualofstyle.org/book/ed17/part3/ch14/psec011.html>).

found that agencies are still not consistently incorporating records management practices and agency records officers into all aspects of the data's life cycle, impacting the future preservation and usability of agency information.¹⁴³ At present, it may also be ambiguous as to whether and when information has been altered by an agency and which agency officials were involved in maintaining or modifying the data. Applying the concept of data governance to agency data creation can begin to mitigate against these risks.

In particular, NIST has worked to define *data governance*—or information resources management in a data-specific context—as a set of processes that ensures that data assets are formally managed throughout an organization and that such processes establish authority and decisionmaking parameters related to the data.¹⁴⁴ Related to this effort, in 2024 NIST began work on a new project seeking to integrate existing frameworks related to privacy, cybersecurity, and artificial intelligence to support a data governance framework that could support data quality, data value, and accountability objectives.¹⁴⁵

Although Congress could work to ensure that agencies are properly applying such roles and responsibilities within their data producing programs, another way of inspecting the quality and integrity of federal data is to continue efforts to make such data transparently available to the public. Because of the quantity of federal data being produced, leveraging community interest in specific datasets could be another way to strengthen accountability by documenting organizational data decisions through additional public disclosure requirements.

¹⁴³ NARA, *System Inspection (Multi-Agency Report): Structured Data Managed Within Databases*, p. 6.

¹⁴⁴ NIST, Computer Security Resource Center, “Glossary: Data Governance,” https://csrc.nist.gov/glossary/term/data_governance.

¹⁴⁵ NIST, “Data Governance and Management (DGM) Profile,” December 11, 2024, <https://www.nist.gov/privacy-framework/new-projects/data-governance-and-management-profile>. For information on cybersecurity, see also CRS In Focus IF10559, *Cybersecurity: A Primer*, by Chris Jaikaran. Systems security and cybersecurity are outside the scope of this report.

Appendix. Glossary

Term	Definition
Availability	The timely and reliable access to and use of information
Comprehensive data inventory (CDI)	Accounts for all data assets created by, collected by, under the control or direction of, or maintained by an agency
Data	For purposes of this report, the type of structured and semi-structured information stored in a standardized format
Data asset	A collection of data elements or datasets that may be grouped together that is composed of structured or semi-structured data
Data governance	A set of processes that ensures that data assets are formally managed throughout the enterprise; a data governance model establishes authority and management and decisionmaking parameters related to the data produced or managed by the enterprise
Federal record	Consists of all recorded information, regardless of form or characteristics, made or received by a federal agency under federal law or in connection with the transaction of public business and preserved or appropriate for preservation by that agency or its legitimate successor as evidence of the organization, functions, policies, decisions, procedures, operations, or other activities of the U.S. government or because of the informational value of data in them but excludes library and museum material made or acquired and preserved solely for reference or exhibition purposes and/or duplicate copies of records preserved only for convenience
Fixity	The assurance that a digital file has remained unchanged
Federal Statistical Research Data Center (FSRDC)	Established by the ICSP to provide researchers with access to certain restricted-use statistics produced by RSAUs
General records schedule (GRS)	Issued by NARA and authorizes, after specified periods of time, the destruction of temporary records or the transfer of permanent records to the Archives that are common to several or all agencies
Information collection request (ICR)	Agencies submit ICRs to OMB via OIRA to create, renew, modify, or discontinue information collections
Interagency Council on Statistical Policy (ICSP)	Helps coordinate statistical activities across the federal government
Information	Any communication or representation of knowledge such as facts, data, or opinions in any medium or form, including textual, numerical, graphic, cartographic, narrative, or audiovisual forms
Information life cycle	The stages through which information passes, typically characterized as creation or collection, processing, dissemination, use, storage, and disposition—to include destruction and deletion
Metadata	Structural or descriptive information about data such as content, format, source, rights, accuracy, provenance, frequency, periodicity, granularity, publisher or responsible party, contact information, method of collection, and other descriptions
National Archives and Records Administration (NARA)	An agency that coordinates the federal government's efforts to identify, manage, and preserve records materials
Office of Information and Regulatory Affairs (OIRA)	The agency within OMB that has roles related to the review of regulations and information collections, information policy, and statistical policy

Provenance	The comprehensive documentation of a dataset's origin, transformation history, ownership chain, access controls, and usage patterns across its life cycle
Principal statistical agency (PSA)	An agency whose guiding mission is to produce statistics
Recognized statistical agency (RSAU)	Agencies and units include PSAs and smaller units within agencies that produce statistics
Standard application process (SAP)	The process by which researchers are able to request information for specified projects
Structured data	A physical data model that describes in detail how the data are to be represented and how a representation should be interpreted; may be found in a database or other mechanism that clearly indicates what type of information each data field contains, such as customer ID or part number
Statistical purpose	The description, estimation, or analysis of the characteristics of groups without identifying the individuals or organizations that comprise such group and the development, implementation, or maintenance of methods, technical or administrative procedures, or information resources that support those activities
Unstructured data	Does not follow a detailed data model or format and often lacks an explicit structure, such as documents, pictures, audio, and video

Author Information

Meghan M. Stuessy
Analyst in Government Organization and
Management

Taylor R. Knoedl
Analyst in American National Government

Disclaimer

This document was prepared by the Congressional Research Service (CRS). CRS serves as nonpartisan shared staff to congressional committees and Members of Congress. It operates solely at the behest of and under the direction of Congress. Information in a CRS Report should not be relied upon for purposes other than public understanding of information that has been provided by CRS to Members of Congress in connection with CRS's institutional role. CRS Reports, as a work of the United States Government, are not subject to copyright protection in the United States. Any CRS Report may be reproduced and distributed in its entirety without permission from CRS. However, as a CRS Report may include copyrighted images or material from a third party, you may need to obtain the permission of the copyright holder if you wish to copy or otherwise use copyrighted material.